5-2018

A Recency-Only Pareto/NBD

Rohan Rajagopalan
University of Pennsylvania

A Recency-Only Pareto/NBD

Abstract
This paper will examine an extension to the Pareto/NBD model commonly used in customer base analysis. The paper will demonstrate and estimate a "recency-only" version of the Pareto/NBD. The model will be fit to both real-world and simulated data and will be compared to the "recency/frequency" model. The paper finds that the model can adequately recover parameters in a variety of settings, but in some cases may predict poorly.

Keywords
pareto/NBD, customer-base analysis, recency/frequency, survey, telemetry

Disciplines
Business | Marketing

A Recency-Only Pareto/NBD

By

Rohan Rajagopalan

An Undergraduate Thesis submitted in partial fulfillment of the requirements for the JOSEPH WHARTON SCHOLARS program

Faculty Advisor:

Peter S. Fader

Frances and Pei-Yuan Chia Professor

THE WHARTON SCHOOL, UNIVERSITY OF PENNSYLVANIA

MAY 2018

Contents

1 Abstract

This paper will examine an extension to the Pareto/NBD model commonly used in customer base analysis. The paper will demonstrate and estimate a "recency-only" version of the Pareto/NBD. The model will be fit to both real-world and simulated data and will be compared to the "recency/frequency" model. The paper finds that the model can adequately recover parameters in a variety of settings, but in some cases may predict poorly.

Keywords: *Pareto/NBD, Customer-base analysis, recency/frequency, survey, telemetry*

2 Introduction

This thesis will examine an adaptation of the standard Pareto/NBD model. Where the traditional model is estimated using the sufficient information of recency and frequency, the proposed model will only require recency. This extension has many potential implications. A recency-only model could form the backbone for survey-based CLV calculations or resolve the common problem of left-censored customer data.

Section 3 discusses existing literature surrounding customer base analysis, the Pareto/NBD and censored data. Section 4 presents the recency-only model. Section 5 discusses the validation approach and evaluative metrics. The model is validated on the canonical CD-Now data set and via simulated purchasing "worlds." The last sections will discuss the results and offer possible extensions.

3 A Review of Existing Literature

3.1 Customer Base Analysis

Customer base analysis is a topic of managerial importance (Gordon 1982). Customer base analysis can answer managerial questions surrounding customer activity and future transactions. The Pareto/NBD has been frequently used in customer base analysis (Schmittlein, Morrison and Colombo 1987). The Pareto/NBD provides a mathematical model for customer churn and purchase frequency (Schmittlein et. al. 1987). Research in the area has, to this point, focused on modeling customer behavior based on both frequency and recency. That is, the model requires knowledge of how many purchases a customer has made within a certain period and when the customer most recently purchased. The proposed recency-only model will only require knowledge of customer purchase recency.

Research in customer base analysis has shown that managers can derive true insight into their business by understanding their customers. Customer base analysis can reveal how loyalty programs affect profitability. Analysis in some settings has found a positive relationship between customer loyalty and customer profitability (Reinartz and Kumar 2000). Customer base analysis can also reveal how customer purchasing impacts churn likelihood. One study found unintuitive dynamics between purchase frequency and duration of customer lifetime (Reinartz and Kumar 2003). Furthermore, customer base analysis provides support for targeted relationship marketing efforts. Research has found that customers who renew are more satisfied than those who churn (Crosby and Stephens 1987). Furthermore, longer tenured customers tend to be more satisfied than recent acquisitions (Bolton 1998). Understanding customer churn likelihoods can give the manager time to intervene before the customer churns (Allenby, Leone and Jen 1999). Having a robust model for customer base analysis like the Pareto/NBD gives the manager an important tool that can be used to derive a variety of insights.

3.2 The Pareto/NBD

First proposed by Schmittlein et. al. (1987), The Pareto/NBD has been highly successful as a tool for customer base analysis. The Pareto/NBD aims to model whether or not customers are alive and, if alive, how frequently they purchase. Customers purchase according to a Poisson process while alive. Customer lifetimes are distributed according to an exponential distribution. Purchasing rates and survival propensities vary across the population according to separate gamma distributions (Schmittlein et. al. 1987).

Since 1987, the Pareto/NBD has been extended by other researchers. The Pareto/NBD and other similar models can be used to solve a number of managerial problems including estimating the number of "active" customers, ranking customers based on probability of being “alive,” and predicting future transaction levels. The model has been shown to work well in a range of settings (Schmittlein and Peterson 1994). Fader, Hardie and Lee have vaidated the model’s forward-looking predictions for an online music retailer (2005a). Abe has applied the model in different purchasing settings from e-commerce to department stores to large-scale chains (2009).

In addition to empirical validation, the model has been applied in several other areas. Hopmann and Thede used it to investigate churn forecasts in non-contractual settings (2005). Wübben and Wangenheim showed the model performs equivalently or better than common managerial heuristics (2008). Glady, Baesens and Croux extend the model to provide estimates of customer lifetime value in several settings (2009). In these varied settings the Pareto/NBD has formed the backbone of wide-ranging and successful customer base analysis.

3.3 Censored Data

A modified Pareto/NBD would enable customer base analysis in business settings where frequency information is unknown, inaccurate, or hard to obtain. There has been limited research on fitting the model in these kinds of settings where data is censored. Research that has addressed censoring issues has focused on interval censoring (Fader and Hardie 2010).

It seems plausible that in some business settings, customer purchase frequency has not been recorded. An adaptation of the Pareto/NBD that can account for this missing information will allow managers in these settings to gain the many insights provided by customer-base analysis. If frequency information has not been recorded, managers could survey their customer base and implement survey-based CLV. Research has shown long-term recency effects in human memory suggesting customers may accurately remember the time of their most recent purchase (Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, Usher 2005). Furthermore, research has shown frequency is poorly remembered over longer time frames in some settings (Ritter et. al. 2001). A recency-only Pareto/NBD would take advantage of these inherent traits and help make survey-based CLV possible.

4 Recency-Only Pareto/NBD

4.1 The Standard Model

This section will briefly discuss the mathematical intuition behind the standard Pareto/NBD (RF) and the the proposed recency-only (RO) model.

As discussed, the standard Pareto/NBD describes two independent processes (Schmittlein et. al. 1987). First, a customer's lifetime τ is modeled as an exponential process with "death" rate μ:

$$f(\tau|\mu) = \mu e^{-\mu\tau} \tag{1}$$

Given that the customer is alive until time τ, purchase frequency is modeled with a Poisson distribution with purchasing rate λ:

$$P(X = x|\lambda, \tau > T) = e^{-\lambda T}\frac{(\lambda T)^x}{x!} \qquad\qquad x = 0, 1, 2, \ldots \tag{2}$$

To capture heterogeneity across the population, it is assumed that μ is distributed according to a gamma distribution with shape parameter s and scale parameter β. Similarly heterogeneity on λ is distributed according to a gamma distribution with shape parameter r and scale parameter α.

The Pareto/NBD has a general likelihood function of

$$L(r, \alpha, s, \beta|x, t_x, T) = \frac{\lambda^x \mu}{\lambda + \mu}e^{-(\lambda+\mu)t_x} + \frac{\lambda^{x+1}}{\lambda + \mu}e^{-(\lambda+\mu)T} \tag{3}$$

where x is the number of observed purchases in the interval (0, T], and t_x is the time of the most recent purchase (Fader and Hardie 2010). Estimations of the Pareto/NBD require knowledge of x and t_x (and T).

4.2 The Recency-Only Pareto/NBD

This section presents a Recency-Only version of the same model. Estimations of the model will only require knowledge of t_x (and T). This model was first presented by Paramveer Dhillon (2016).[1]

The model considers two main cases:

1. The customer made 0 purchases in the interval $(0, T]$ and thus $t_x = 0$
2. The customer made $x > 0$ purchases with the most recent purchase occurring at time $0 \leq t_x \leq T$

Each case leads to two further sub-cases.

<u>Case 1:</u>

1. The customer made 0 purchases in the interval and is alive at time T
2. The customer made 0 purchases in the interval and died at some time τ where $0 \leq \tau \leq T$

This gives the following likelihood

$$L(\lambda, \mu | t_x > 0, T) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)T} \tag{4}$$

[1]The full derivation as presented by Dhillon is available in the appendix.

Case 2:

1. The customer made his or her last purchase at time t_x and is alive at time T
2. The customer made his or her last purchase at time t_x and died at some time τ where $t_x \leq \tau \leq T$

This gives the following likelihood

$$L(\lambda, \mu | t_x > 0, T) = \frac{\lambda\mu}{\lambda + \mu} e^{-\mu t_x} + \frac{\lambda^2}{\lambda + \mu} e^{-(\lambda+\mu)T} e^{\lambda t_x} \tag{5}$$

Conditioning on λ and μ can be removed via integration as with the Pareto/NBD. Let

$$L(r, \alpha, s, \beta | t_x, T) = A_1 + A_2 \tag{6}$$

$$L(r, \alpha, s, \beta | t_x > 0, T) = A_3 + A_4 \tag{7}$$

This gives a total likelihood of

$$L(r, \alpha, s, \beta) = (A_1 + A_2)^{nZ} + (A_3 + A_4)^{nNZ} \tag{8}$$

where nZ is the number of customers for whom $t_x = 0$ and nNZ is the number for whom $t_x > 0$.

5 Data and Methodology

The following sections will discuss the validation of the model. In general, the validation approach will be to fit the model to data and to compare the parameters and forecasts of the recency-only (RO) model to those of the traditional Pareto/NBD (RF) or the "true" simulated parameters. The RO model can be considered to perform well if its forecasts and parameters are similar to those of the RF model, as the Pareto/NBD has been empirically validated many times.

5.1 CD-Now

The CD-NOW dataset is commonly used in the Pareto/NBD literature (Fader and Hardie 2005; Abe 2009). It provides weekly customer purchase data for an online music retailer from January 1997 until June 1998. The data set is a typical example of a real-life purchasing environment.

The RO and RF models are fit to this data set and parameter estimates are compared. The performance is also measured via traditional metrics. The in-sample histogram, out-of-sample histogram, incremental tracking plot and cumulative tracking plot are presented.

5.2 81 Worlds

One of the drawbacks of validation through the CD-NOW data set is that the parameter estimates can only be compared to another imperfect modeling approach. The recovered parameters are never compared to the "true" underlying parameters because those parameters are unknown. Simulation presents a solution to this problem. By generating data and fitting both models to that data, the recovered parameters can be compared to the "true" values across a variety of purchasing settings.

Data is simulated following the same procedure as Jerath, Fader and Hardie (2013). Data is generated according to the Pareto/NBD process described in section 4. Using the four parameters of the Pareto/NBD, weekly customer data is generated for 81 “parameter-worlds” at 3 values of each parameter. The values are shown below:

$$r, s = (0.5, 1, 1.5)$$

$$\alpha, \beta = (5, 10, 15)$$

These values capture a wide range of purchasing settings from low-purchasing, high-death worlds to high-purchasing, low-death worlds. This gives an opportunity to see if the RO model is more stable and fits better under certain purchasing circumstances. Data is generated for 10000 customers for 104 weeks. Data on both recency and frequency is generated to fit both the RF and RO models. Each "world" is simulated 10 times to account for the randomness inherent in the data generation process.

The RO and RF models are fit to the first 52 weeks of the generated data. The second 52 week period is used as a holdout period. Parameter RMSEs are computed for both models. The out-of-sample fit is measured by χ^2 goodness-of-fit statistics for predicted holdout transactions and by the mean absolute percent error of Discount Expected Transactions[2]

The data generation method used by Jerath et. al. and replicated here has the advantage of creating a variety of purchasing settings. As in Jerath et. al., additional metrics are presented for five representative worlds: World MM, World LL, World LH, World HL, and World HH. These worlds differ in the underlying parameter values used to generate the data.

$$\text{World MM: } r = 0.5, \alpha = 5, s = 0.5, \beta = 5$$

[2]Discount Expected Transactions or DET is a managerially relevant measure of out-of-sample performance that is closely related to Customer Lifetime Value. DET is a prediction based on the recovered parameters and the observed purchasing history of a given customer. DET represents the predicted number of future transactions discounted to present day. If a customer spend is modeled, DET can easily be used to compute CLV.

World LL: $r = 0.5, \alpha = 15, s = 1.5, \beta = 5$

World LH: $r = 0.5, \alpha = 15, s = 0.5, \beta = 15$

World HL: $r = 1.5, \alpha = 5, s = 1.5, \beta = 5$

World HH: $r = 1.5, \alpha = 5, s = 0.5, \beta = 15$

These specific worlds were chosen to show different purchasing cases. A robust model should fit well in each of these varied settings (Jerath et. al. 2013). For these worlds, out-of-sample fit will be measured by χ^2 goodness-of-fit statistics on the expected and actual number of holdout-period transactions. In-sample-fit will be measured by the "likelihood distortion metric" used by Jerath et. al. (2013).

Distortion Metric

As the models are not fit to the same data, their log-likelihoods are not comparable. Instead, the following approach is used. The parameters are estimated for both the RF and RO models. The parameters recovered for the RO model are used to calculate a log-likelihood based on both the recency and frequency data.

The distortion metric D_i for world i is then calculated as follows.

$$D_i = \left| \frac{ll_{(RF,\ i)} - ll_{(RO,\ i)}}{ll_{(RF,\ i)}} \right| \times 100$$

where $ll_{(RO,\ i)}$ is the log-likelihood from the RF model for the ith world using the parameters recovered from the RO model.

Examining the formula shows that it is bounded between 0 and 100. Values closer to 0 suggest that the RO model has not lost much information relative to the RF model.

6 Results

6.1 CD-Now

The first section presents the results from fitting both the RF and RO models to the CD-Now data.

Parameter Estimates

The parameter estimates capture the underlying distributions that govern the model. r and α describe the purchasing process while s and β describe the death process.

	r	$\boldsymbol{\alpha}$	**s**	$\boldsymbol{\beta}$	$\boldsymbol{E[\lambda]}$	$\boldsymbol{E[\mu]}$
Recency-Frequency	0.6519	11.7951	0.5391	8.0669	0.0553	0.0668
Recency-Only	0.5533	10.5775	0.6062	11.6682	0.0523	0.0520

The results of the CD-NOW parameter estimates for both the recency-frequency (RF) and recency-only (RO) models are reported above. For the CD-NOW dataset, the parameter estimates track well. The expected means for λ and μ are also similar. The recency-only model appears to have slightly lower "purchasing" but also has slightly lower "dying."

In-Sample Histogram

The in-sample histogram shows the quality of fit within the portion of the data used to fit the model. In some instances the parameters may not be identical, but the in-sample histogram can still look quite good. For instance, lower purchasing may "cancel out" lower dying.

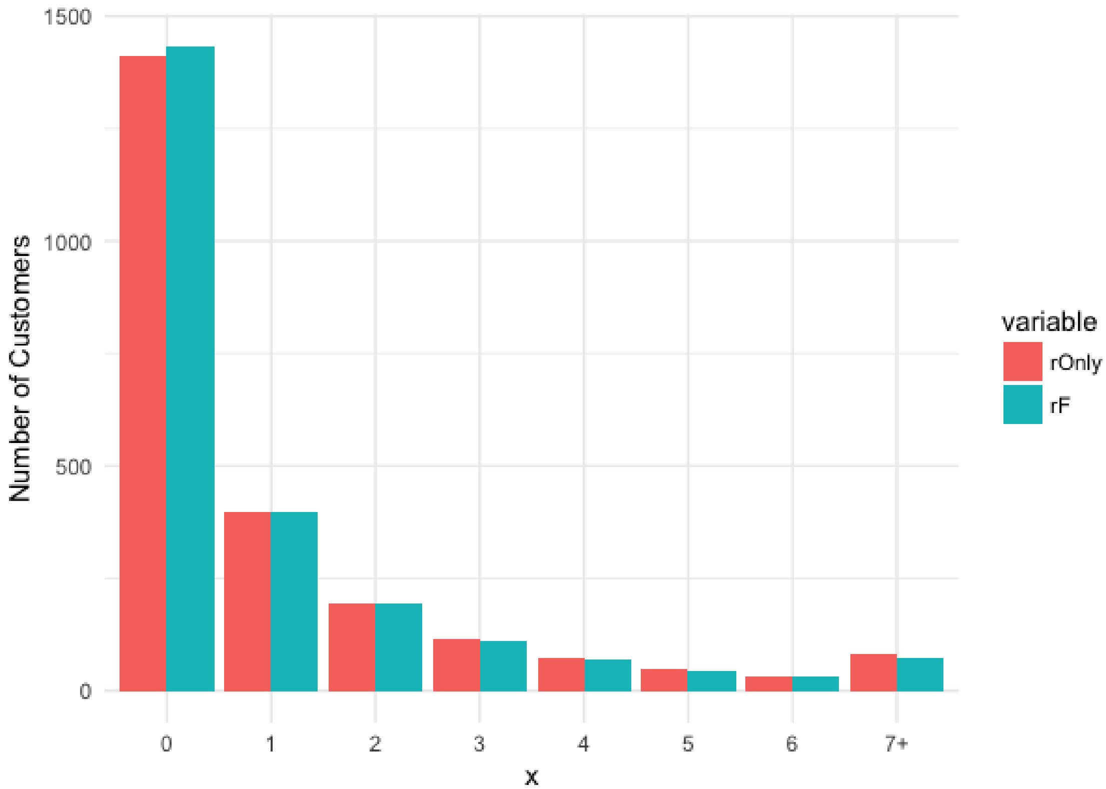

Figure 1: Expected number of customers making x transactions in weeks 0-39. Reported for the RO and RF models.

The model was fit to data from weeks 0-39. The χ^2 statistic is 2.0955. The closer to 0 the χ^2 statistic, the more similar the two histograms. It can be seen, both visually and from the statistic that both models have very similar in-sample predictions. These predictions can also be compared to the observed data.

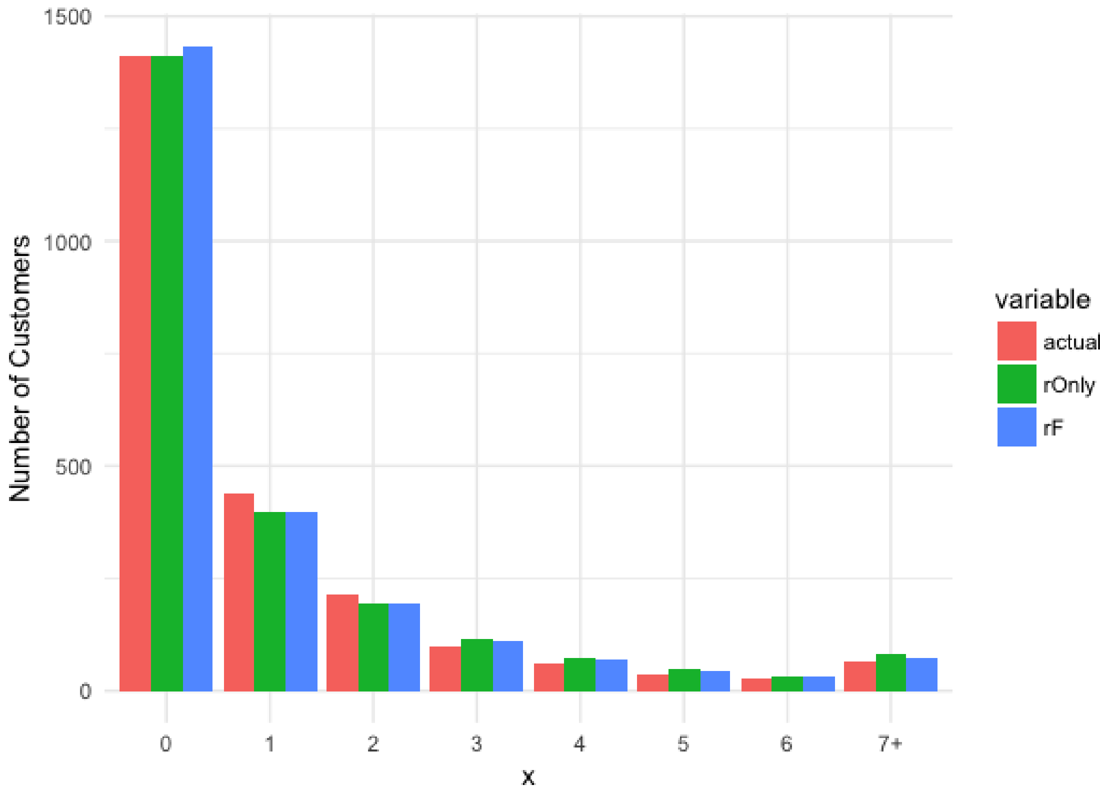

Figure 2: Expected number of customers making x transactions in weeks 0-39. Reported for the observed data and the RO and RF expectations.

Here we see both models track fairly well. The RO model performs better for $x = 0$ but worse for $x > 3$. The χ^2 statistic between the RO-expected and the observed data is 16.6734. This is in comparison to a χ^2 statistic of 11.9923 between R-F and the observed data. For the CD-Now data set the RO model fits adequately in-sample.

Out of Sample Histogram

Managers are often more interested in out-of-sample predictions. For the RO model to be of managerial use, it must forecast accurately. The out-of-sample histogram shows the expected number of customers making $x = 0, 1, 2, \ldots$ transactions in the interval $[T, T + T^*]$.

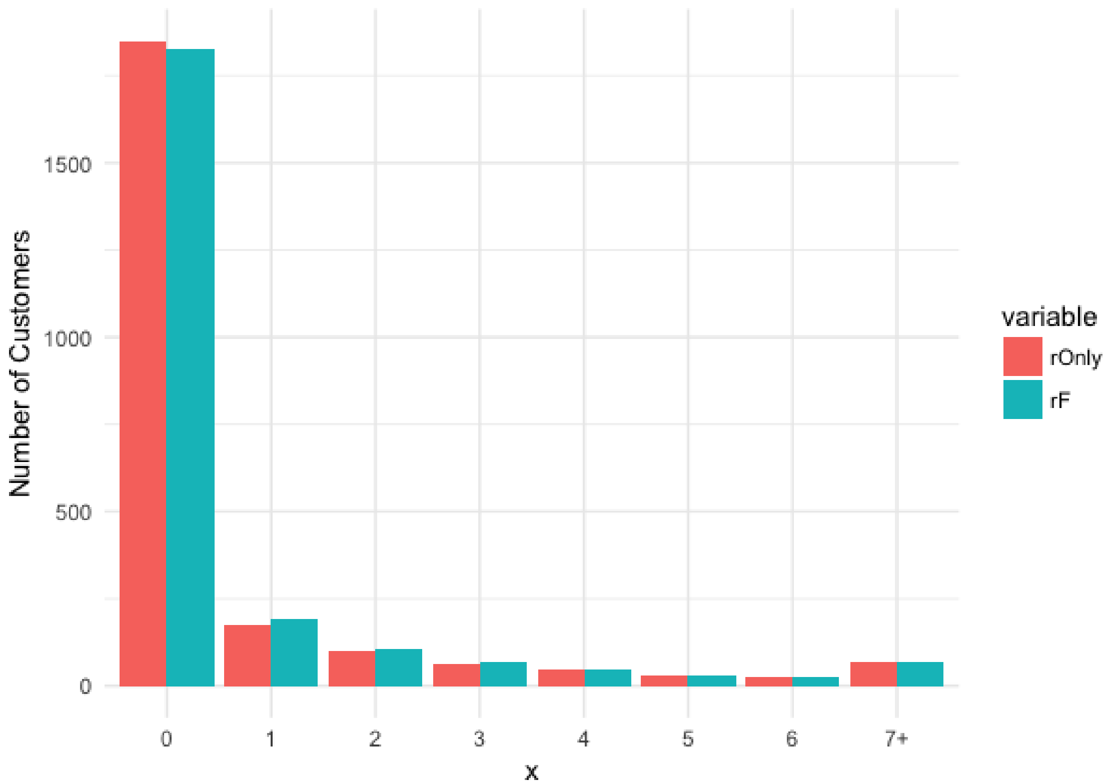

Figure 3: Expected number of customers making x transactions in weeks 39-78. Reported for the RO and RF models.

The out-of-sample histogram shows that both models forecast a similar number of transactions. The χ^2 statistic between the two models is 2.3233. Again this is very close to 0, suggesting the forecasts are similar. These forecasts can also be compared to the observed number of transactions.

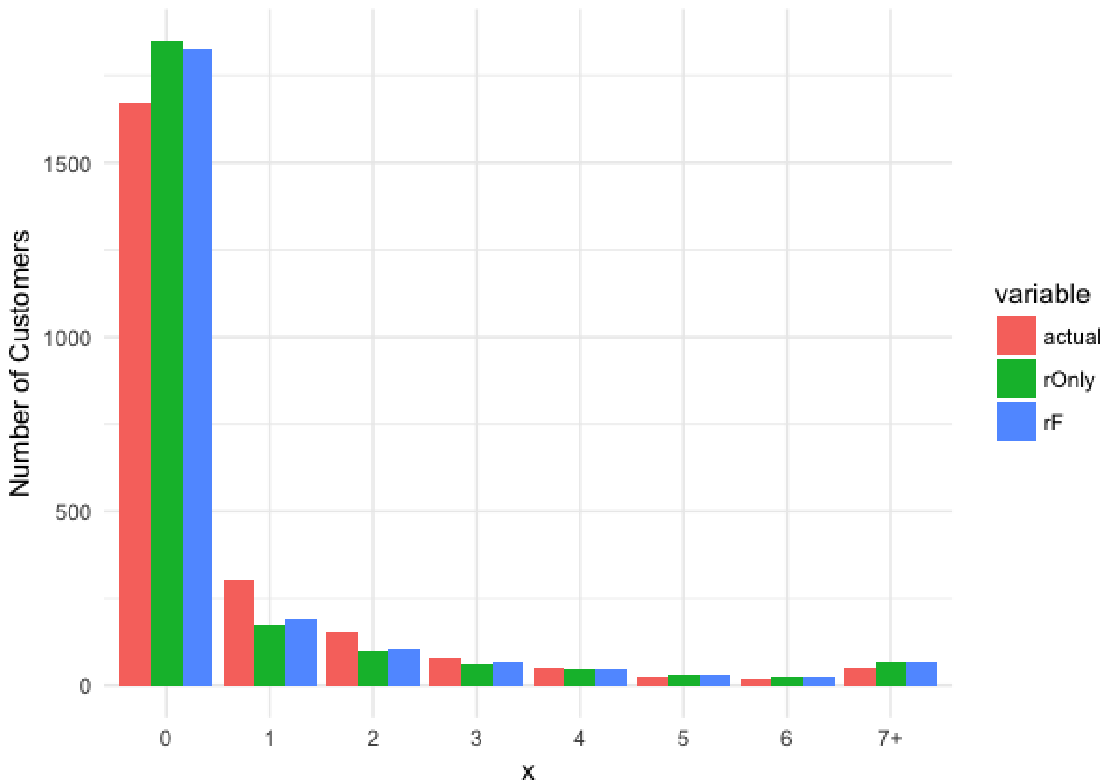

Figure 4: Expected number of customers making x transactions in weeks 39-78. Reported for the observed data and the RO and RF models.

The out-of-sample predictions are plotted against the observed data from the holdout period. The holdout period for this model was weeks 39-78. There is some lack of fit between both models and the actual data. The RO model fit is worse than that of the RF model. This is to be expected given the loss of information. The χ^2 statistic between the RO model and the actual data is 148.8003. In comparison, the χ^2 statistic between the RF model and actual data is 107.8927. The increased lack-of-fit in the holdout period suggests that predictions using the RO model may be worse than those of the traditional Pareto/NBD.

Incremental Tracking

The incremental tracking plot shows the number of incremental repeat transactions made each week. The actual data is quite noisy, but a good model will fit the general curve. This plot can be used to asses fit both in and out-of-sample.

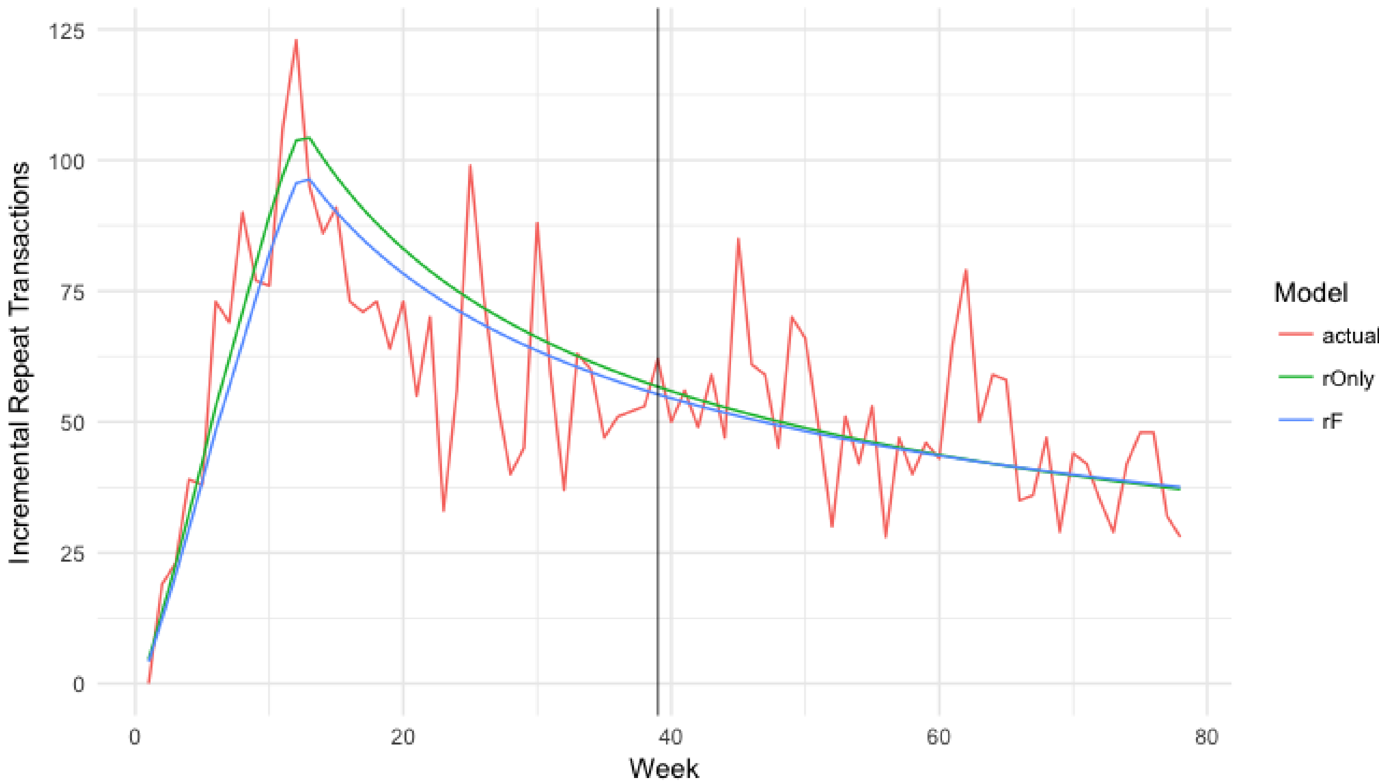

Figure 5: Incremental repeat transactions for weeks 0-78. Reported for the observed data and both models

The incremental tracking plot looks very similar for both models. There is considerable noise in the data but both models seem to capture the general trend. The RO model predicts more incremental transactions before the holdout period.

Cumulative Tracking

The cumulative tracking plot is similar to the incremental plot. It can be viewed as summing the number of incremental transactions over the rolling window from weeks 0-78. The plot also shows both in and out-of-sample fit.

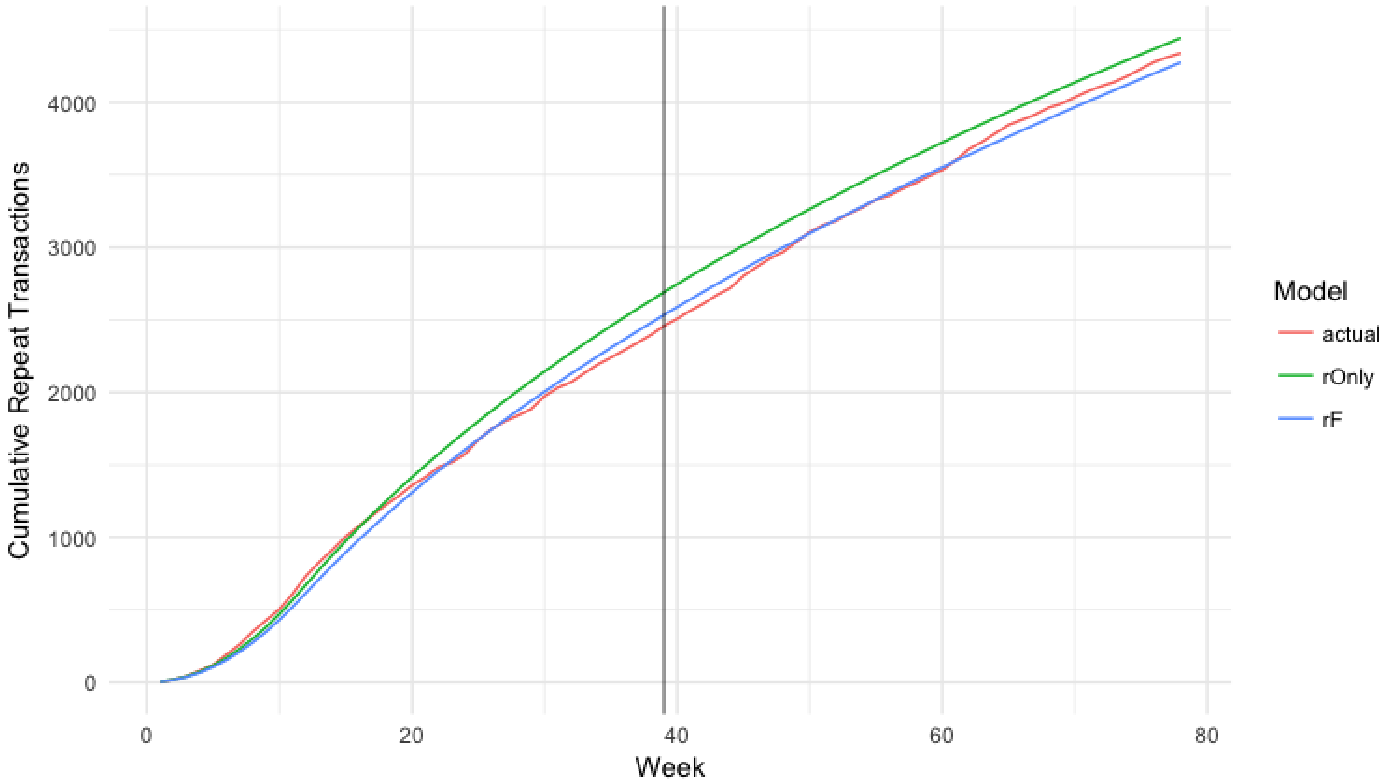

Figure 6: Cumulative repeat transactions for weeks 0-78. Reported for the observed data and the RO and RF models.

The cumulative tracking plot shows that the RO model tends to over-predict the number of transactions. While initially all three lines track together, by week 20 the RO model tends to predict more cumulative transactions. This supports the increased lack-of-fit suggested by the out-of-sample χ^2. It should be noted that while there is some lack-of-fit, the absolute size of the error is at most 200-300 transactions. The observed cumulative number of transactions ends up somewhere between the two models. Given the loss of information, the RO forecasts might still be tolerable.

6.2 81-Worlds

This section discusses the results from the 81 simulated worlds. Each world was simulated 10 times[3] Results across all 81 worlds are reported first. Then specific results for the 5 featured worlds are reported.

Median Parameter RMSE

RMSE is a measure of error between two numbers. In this case the RMSE gives $\sqrt{(p_r - p_t)^2}$ where p_r is a recovered parameter and p_t is the true value used to simulate the data. The median RMSE for each parameter across the 81 worlds is reported.

	r	$\boldsymbol{\alpha}$	**s**	$\boldsymbol{\beta}$
Recency-Frequency	0.001221011	0.1562924	0.001233512	0.5808081
Recency-Only	0.02748866	5.783653	0.001445147	0.7478821

The errors for the "death" process are comparable for both models. The RO model errors are larger but are of the same magnitude as the RF model. The RO model RMSEs for the purchasing process are much higher than that of the RF model. In particular the α parameter has quite large errors.

χ^2 statistic for out-of-sample predictions

As discussed, the errors in the parameters may not totally eliminate the ability of the model to forecast accurately. The χ^2 statistics for the holdout period are summarized below.

	Mean	**Min**	**1st Quartile**	**Median**	**3rd Quartile**	**Max**
Recency-Frequency	6.309	3.990	5.313	6.184	7.191	9.276
Recency-Only	8.967	4.417	7.081	8.237	10.129	25.700

[3]The specific worlds noted in section 5 were simulated an 5 times each.

The mean and median value across all 81 world is fairly low. The RO model performs worse, but not by a considerable margin. The RO distribution of χ^2 statistics has a longer tail with some high values. This suggests in some circumstances the predictions may not be accurate.

Mean Absolute Percent Error (MAPE) of Discounted Expected Transactions (DET)

Another measure of out-of-sample prediction is DET. DET is a forward-looking metric that is useful for computing CLV.[4] The mean absolute percent error on this metric across the 81 worlds is summarized below. The summary is computed using the median value for the 10 runs.

	Mean	Min	1st Quartile	Median	3rd Quartile	Max
Recency-Frequency	9.707	4.618	7.505	9.402	11.322	21.237
Recency-Only	15.160	5.133	10.278	14.334	19.031	28.854

The MAPE on DET shows similar patterns to that of the χ^2. The RO model performs worse than the RF model. The average MAPE is about 50 percent larger. There is less variation at the maximum value than with the χ^2 statistic.

6.2.1 Featured Worlds

The following section reports an in-sample and out-of-sample metric for the five feature worlds. The metrics are computed in 5 runs of each world.

Likelihood Distortion Metric (%)

The likelihood distortion metric described in section 5 is tabulated below. Intuitively, this metric shows how much information is "lost" when using the RO parameters on the recency and frequency data. Values closer to 0 suggest not much information has been lost.

[4]See section 5.

MM (%)	LL (%)	LH (%)	HL (%)	HH (%)
0.00214871	0.4660355	0.0008462963	0.05726152	0.002586775

Across the five worlds the distortion metric is fairly good. In most of the worlds the value is less than 0.05%. World LL has a higher value suggesting there may be some inability to accurately fit the RO model in that purchasing setting.

Out of sample χ^2

MM	LL	LH	HL	HH
6.598902	9.316768	4.313535	5.198987	10.14952
6.740242	11.320646	4.387782	4.710081	11.00484

The out-of-sample χ^2 shows similar results. The values are fairly close for most of the worlds. In world LL, the RO model has higher out-of-sample error which supports the lack of fit suggested by the distortion metric.

7 Discussion

7.0.1 Discussion of Results

The results shown in section 6 demonstrate two things. First, the recency-only model can often adequately recover parameter estimates. Given that the RO model has "thrown away" half of the data available to the RF model, it is quite remarkable that it is able to accurate parameter estimates. Second, while these estimates are often quite good (particularly given the loss of data), the RO model does not forecast or fit as well as the RF model and in some cases is significantly worse. In particular, the errors on the r and α parameters are significantly higher than those of the RF model in the 81-worlds simulations.

The practical usefulness of the model will depend on the circumstances in which it is being used. If there is truly no frequency information, the RO model will, at the very least, provide some way of conducting customer-base analysis. However, the fit and forecast may not be as robust as is typically seen with the RF model. More work must be done on determining the purchasing settings in which parameter estimates and forecasts are stable and accurate.

7.0.2 Extensions

Left Censoring

In circumstances where a firm has not recorded customer purchase information for some time and then begins to do so, the data is left-censored. Once the firm starts recording purchases, the first recorded purchase of Customer A could in fact be his or her 15th over his or her customer-lifetime. Contrastingly, the first recorded purchase of Customer B might actually be his or her first. This creates problems with estimating CLV, as pre-observation frequencies are not known.

The RO model may provide a solution to this problem. If the customers are observed for some time period, the time of the most recent purchase will be known. Thus, instead of worrying about the missing frequency information the RO model could be used to perform customer-base-analysis. This would still require some way to capture T_n or the customer's total lifetime with the firm.

Survey CLV

This idea can be extended to "Survey-Based CLV." A firm with no customer purchasing information could conduct a survey of its customers as to their most recent time of purchase and the time of their first purchase. This would provide the sufficient information of T_n and t_x required to estimate the RO model. More work must be done to assess the stability of parameter estimates in this scenario particularly given the noise in survey responses.

In addition, this method would likely also require an interval censored extension of this model. Customers would almost certainly be unable to remember their most recent and first times of purchase down to a specific day. It is more feasible that they could remember a specific interval in which they had purchased. This is discussed further in a subsequent section.

Stability of Parameters

In typical customer-base analysis, the model is fit to some calibration period. With the RF model, the length of this period is unlikely to impact the parameter estimates. Shifting this period will change recency but will leave frequency more or less stable. However, for the RO model changing recency may have a larger impact on the stability of parameter estimates. Fitting the model to calibration periods of different length and examining the parameter estimates will provide some sense of how much this impacts the model.

Additionally, the RO model parameters can vary significantly from estimation to estimation even if the calibration period remains the same. This may be more common in some purchasing settings. Determining what kinds of underlying data cause this to happen and how many estimations are needed before parameters stabilize will improve managerial confidence in using the model in real-world settings.

Interval-Censoring

Fader and Hardie have presented an interval-censored extension to the Pareto/NBD (2005b). They provide a likelihood function for data where frequency is known but recency is interval censored. A similar expression can be derived for the RO model (Dhillon 2016). However, the stability of this model is unknown. Further simulation analysis could show the viability of this method and increase the possibility of practical survey-based CLV.

8 References

Abe, Makoto. 2009. "'Counting Your Customers' One by One: A Hierarchical Bayes Extension to the Pareto/NBD Model." *Marketing Science* 28 (3): 541–53.

Akaike, Hirotugu. 1987. "Factor Analysis and AIC." In *Selected Papers of Hirotugu Akaike*, 371–86. Springer Series in Statistics. Springer, New York, NY.

Allenby, Greg M., Robert P. Leone, and Lichung Jen. 1999. "A Dynamic Model of Purchase Timing with Application to Direct Marketing." *Journal of the American Statistical Association* 94 (446): 365–74.

Bolton, Ruth N. 1998. "A Dynamic Model of the Duration of the Customer's Relationship with a Continuous Service Provider: The Role of Satisfaction." *Marketing Science* 17 (1): 45–65.

Crosby, Lawrence A., and Nancy Stephens. 1987. "Effects of Relationship Marketing on Satisfaction, Retention, and Prices in the Life Insurance Industry." *Journal of Marketing* Research 24 (4): 404.

Davelaar, Eddy J., Yonatan Goshen-Gottstein, Amir Ashkenazi, Henk J. Haarmann, and Marius Usher. 2005. "The Demise of Short-Term Memory Revisited: Empirical and Computational Investigations of Recency Effects." *Psychological Review* 112 (1): 3–42.

Dhillon, Paramveer. 06/15. "All You Need Is 'R': Estimating Pareto/NBD Using Only 'R'Ecency Data."

Canning, Gordon. 1982. "Do a Value Analysis of Your Customer Base." *Industrial Marketing Management* 11 (2): 89–93.

Fader, Peter, Bruce Hardie, Yuzhou Liu, Joseph P. Davin, and Thomas J. Steenburgh. 2018. "'How to Project Customer Retention' Revisited: The Role of Duration Dependence." SSRN Scholarly Paper ID 3100346. Rochester, NY: Social Science Research Network.

Fader, Peter S., and Bruce G. S. Hardie. 2005a. “The Value of Simple Models in New Product Forecasting and Customer-Base Analysis.” *Applied Stochastic Models in Business and Industry* 21 (4–5): 461–73.

Fader, Peter S, and Bruce G S Hardie. 2005b. “A Note on Deriving the Pareto/NBD Model and Related Expressions,” 14.

Fader, Peter S, and Bruce G S Hardie. 2010. “Implementing the Pareto/NBD Model Given Interval-Censored Data,” November, 7.

Fader, Peter S., Bruce G.S. Hardie, and Ka Lok Lee. 2005. “RFM and CLV: Using Iso-Value Curves for Customer Base Analysis.” *Journal of Marketing Research* 42 (4): 415–30.

Glady, Nicolas, Bart Baesens, and Christophe Croux. 2009. “Modeling Churn Using Customer Lifetime Value.” *European Journal of Operational Research* 197 (1): 402–11.

Hopmann, Jörg, and Anke Thede. 2005. “Applicability of Customer Churn Forecasts in a Non-Contractual Setting.” In *Innovations in Classification, Data Science, and Information Systems*, 330–37. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Berlin, Heidelberg.

Jerath, Kinshuk, Peter Fader, and Bruce Hardie. 2013. “Customer-Base Analysis on a ‘Data Diet’: Model Inference Using Repeated Cross-Sectional Summary (RCSS) Data.” SSRN Scholarly Paper ID 1708562. Rochester, NY: Social Science Research Network.

Reinartz, Werner J., and V. Kumar. 2000. “On the Profitability of Long-Life Customers in a Noncontractual Setting: An Empirical Investigation and Implications for Marketing.” *Journal of Marketing* 64 (4): 17–35.

Reinartz, Werner J., and V. Kumar. 2003. “The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration.” *Journal of Marketing* 67 (1): 77–99.

Ritter, P. L., A. L. Stewart, H. Kaymaz, D. S. Sobel, D. A. Block, and K. R. Lorig. 2001.

“Self-Reports of Health Care Utilization Compared to Provider Records.” *Journal of Clinical Epidemiology* 54 (2): 136–41.

Schmittlein, David C., Donald G. Morrison, and Richard Colombo. 1987. “Counting Your Customers: Who Are They and What Will They Do Next?” *Management Science* 33 (1): 1–24.

Schmittlein, David C., and Robert A. Peterson. 1994. “Customer Base Analysis: An Industrial Purchase Process Application.” *Marketing Science* 13 (1): 41–67.

Wübben, Markus, and Florian v Wangenheim. 2008. “Instant Customer Base Analysis: Managerial Heuristics Often ‘Get It Right.’” *Journal of Marketing* 72 (3): 82–93.

9 Appendix

9.1 Recency-Only Derivation

In section 4 the likelihood function for the Recency-Only model was given as:

$$L(r, \alpha, s, \beta) = (A_1 + A_2)^{nZ} + (A_3 + A_4)^{nNZ} \tag{9}$$

This section will show how to compute A_1, A_2, A_3 and A_4. This derivation was compiled and presented by Paramveer Dhillon and is reproduced here (2016).

The first term in equation (1) is the case where a customer has made 0 purchases in the interval $(0, T]$. Computing A_1 and A_2 requires removing the conditioning on λ and μ from the following expression:

$$L(\lambda, \mu | t_x > 0, T) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)T} \tag{10}$$

Removing this conditioning requires the solutions to three separate but related integrals. These solutions are presented by Fader and Hardie (2005b).

The first integral is shown below. ${}_2F_1$ is the Gaussian Hypergeometric Function. The first term in equation (2), A_1, is solved by setting $t = 0, \gamma = 0$ and $\delta = 1$. The second term in equation (2), A_2, is solved by setting $t = T, \gamma = 1$ and $\delta = 0$.

$$\int_0^\infty \int_0^\infty \frac{\gamma^\gamma \mu^\delta}{\gamma + \mu} e^{-(\gamma+\mu)t} g(\gamma|r,\alpha) g(\mu|s,\beta) d\lambda d\mu = \frac{\alpha^r \beta^s}{(\alpha+t)^{r+s+\gamma+\delta-1}} \frac{\Gamma(r+\gamma)\Gamma(s+\delta)}{\Gamma(r)\Gamma(s)} \left(\frac{1}{r+s+\gamma+\delta-1}\right) \times {}_2F_1\left(r+s+\gamma+\delta-1, s+\delta; r+s+\gamma+\delta; \frac{\alpha-\beta}{\alpha+t}\right) \quad (if\ \alpha \geq \beta)$$

$$= \frac{\alpha^r \beta^s}{(\beta+t)^{r+s+\gamma+\delta-1}} \frac{\Gamma(r+\gamma)\Gamma(s+\delta)}{\Gamma(r)\Gamma(s)} \left(\frac{1}{r+s+\gamma+\delta-1}\right) \times {}_2F_1\left(r+s+\gamma+\delta-1, r+\gamma; r+s+\gamma+\delta; \frac{\beta-\alpha}{\beta+t}\right) \quad (if\ \beta \geq \alpha) \tag{11}$$

The second term in equation (1) is the case where a customer has made $x > 0$ purchases in the interval $(0, T]$. The most recent purchase has occurred at time t_x. Computing A_3 and A_4 requires removing the conditioning on λ and μ from the following expression:

$$L(\lambda, \mu | t_x > 0, T) = \frac{\lambda\mu}{\lambda+\mu} e^{-\mu t_x} + \frac{\lambda^2}{\lambda+\mu} e^{-(\lambda+\mu)T} e^{\lambda t_x} \tag{12}$$

This requires the last two integrals. The first term in equation (4), A_3, can be solved by using the third integral with $T = t_x, \gamma = 1$ and $\delta = 1$. The second term in equation (4), A_4, can be solved by using the second integral with $T_1 = T, T_2 = t_x, \gamma = 2$ and $\delta = 0$

$$\begin{aligned}
\int_0^\infty \int_0^\infty \frac{\lambda^\gamma \mu^\delta}{\lambda+\mu} e^{-(\lambda+\mu)T_1} e^{\lambda T_2} g(\lambda|r,\alpha) g(\mu|s,\beta) d\lambda d\mu &= \frac{\alpha^r \beta^s}{(\alpha+T_1)^{r+s+\gamma+\delta-1}} \frac{\Gamma(r+\gamma)\Gamma(s+\delta)}{\Gamma(r)\Gamma(s)} \left(\frac{1}{r+s+\gamma+\delta-1} \right) \\
&\times {}_2F_1\left(r+s+\gamma+\delta-1, s+\delta; r+s+\gamma+\delta; \frac{\alpha-\beta-T_2}{\alpha+T_1} \right) \\
&(if\ \alpha \geq \beta + T_2) \\
&= \frac{\alpha^r \beta^s}{(\beta+T_1+T_2)^{r+s+\gamma+\delta-1}} \frac{\Gamma(r+\gamma)\Gamma(s+\delta)}{\Gamma(r)\Gamma(s)} \left(\frac{1}{r+s+\gamma+\delta-1} \right) \\
&\times {}_2F_1\left(r+s+\gamma+\delta-1, r+\gamma; r+s+\gamma+\delta; \frac{\beta-\alpha+T_2}{\beta+T_1+T_2} \right) \\
&(if\ \beta > \alpha + T_2)
\end{aligned} \tag{13}$$

$$\begin{aligned}
\int_0^\infty \int_0^\infty \frac{\lambda^\gamma \mu^\delta}{\lambda+\mu} e^{-\mu T} g(\lambda|r,\alpha) g(\mu|s,\beta) d\lambda d\mu &= \frac{\alpha^r \beta^s}{(\alpha)^{r+s+\gamma+\delta-1}} \frac{\Gamma(r+\gamma)\Gamma(s+\delta)}{\Gamma(r)\Gamma(s)} \left(\frac{1}{r+s+\gamma+\delta-1} \right) \\
&\times {}_2F_1\left(r+s+\gamma+\delta-1, s+\delta; r+s+\gamma+\delta; \frac{\alpha-\beta-T}{\alpha} \right) \\
&(if\ \alpha \geq \beta + T) \\
&= \frac{\alpha^r \beta^s}{(\beta+T)^{r+s+\gamma+\delta-1}} \frac{\Gamma(r+\gamma)\Gamma(s+\delta)}{\Gamma(r)\Gamma(s)} \left(\frac{1}{r+s+\gamma+\delta-1} \right) \\
&\times {}_2F_1\left(r+s+\gamma+\delta-1, r+\gamma; r+s+\gamma+\delta; \frac{\beta-\alpha+T}{\beta+T} \right) \\
&(if\ \beta > \alpha + T)
\end{aligned} \tag{14}$$

To summarize the results so far:

$$A_1 := T = 0, \gamma = 0, \delta = 1;\ \text{Using equation (3)}$$

$$A_2 := T = T, \gamma = 1, \delta = 0;\ \text{Using equation (3)}$$

$$A_3 := T = t_x, \gamma = 1, \delta = 1;\ \text{Using equation (6)}$$

$$A_4 := T_1 = T, T_2 = t_x, \gamma = 2, \delta = 0;\ \text{Using equation (5)}$$

Substituting these terms into equation (1) will give the solution to $L(r, \alpha, s, \beta)$.